AD-A160 207 ROLE OF AUXILIARY VARIATE AND ADDITIONAL DATA IN   1/1
             DENSITY ESTIMATION(U) PITTSBURGH UNIV PA CENTER FOR
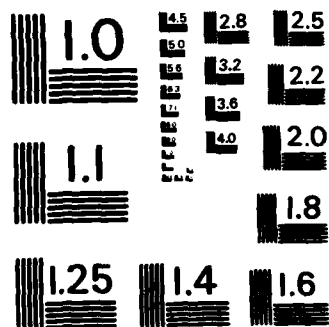   MULTIVARIATE ANALYSIS M AHMAD ET AL. MAY 85 TR-85-21
UNCLASSIFIED AFOSR-TR-85-0748 F49620-85-C-0008   F/G 12/1   NL

END

FILMED

DTIC

MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

ROLE OF AUXILIARY VARIATE AND
ADDITIONAL DATA IN DENSITY ESTIMATION*

by

Manzoor Ahmad**
Université du Québec à Montréal
and
Center for Multivariate Analysis

and

Radhey S. Singh**
University of Guelph

Center for Multivariate Analysis

University of Pittsburgh

DTIC
ELECTE
OCT 1 5 1985
S   D

Approved for public release;
distribution unlimited.

ROLE OF AUXILIARY VARIATE AND
ADDITIONAL DATA IN DENSITY ESTIMATION*

by

Manzoor Ahmad**
Université du Québec à Montréal
and
Center for Multivariate Analysis

and

Radhey S. Singh**
University of Guelph

May 1985

Technical Report No. 85-21

Center for Multivariate Analysis
Fifth Floor, Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

I. INTRODUCTION

Estimation of a density function has drawn considerable attention in the literature over the last two decades. Examples of practical situations calling for the estimation of a density can be found in the works of several authors, e.g., Murthy (1965), Singh (1977b), Liang and Krishnaiah (1985), among others.

In one of the pioneering papers on the problem of non-parametric estimation of a continuous density, a very useful and rather disappointing observation was made by Rosenblatt (1956). According to this observation, any reasonable estimator of a continuous density cannot be unbiased. Therefore, any attempt to improve upon the bias, M.S.E., or rates of convergence involved in the asymptotics, becomes a desirable exercise. The work reported here is an attempt in this direction.

While treating an inference problem relating to a variate Y, a possible approach to gain in precision is to incorporate a concomittant random variable X along with Y. A considerable part of statistical literature has been devoted to this approach. In Section 1 we have proposed some estimators of a univariate probability density function $f(y)$ of a r. v. Y based upon a set of observations taken from a bivariate joint density $\beta(x,y)$ of Y and a suitably chosen concomittant r. v. X, so that $f(y) = \int \beta(x,y)dx$. The estimators have been constructed using some well known heuristic methods employed in some known areas of statistics but never applied in the area of density estimation. Although the asymptotic properties and rates of convergence of these estimators are the same as those of the usual estimator which does not depend on the data on X, we give sufficient

conditions on β and the marginal densities of Y and X under which the proposed estimators would perform better than the usual estimators in the sense of |Bias| and the MSE. The ideas developed here can easily be extended to the case when Y and X are both multivariate.

Different methods of constructing estimators, apparently none better than others in a global sense, (Watson (1969), Wegman (1972a),(1972b)), have appeared in the literature. However, we have adopted the most widely used Rosenblatt (1956) - Parzen (1962) type Kernel method.

In Section 2, we have also looked into the problem of estimating a conditional density $g(y|x)$ of a r.v. Y given another r.v. X based on a set of paired observations on $(X,Y) \sim \beta(x,y)$ and a set of additional observations on $X \sim f(x)$. This problem without the use of additional data has been treated by Rosenblatt (1969). We have obtained better approximation for the variance as compared to Rosenblatt (1969), and have given sufficient conditions on β and f under which the use of additional data on X gives smaller absolute error and variance (and hence the mean squared error) than those obtained without using the additional data. These conditions need to be examined more carefully to ease their accessibility to practical problems.

1. RATIO TYPE KERNEL ESTIMATORS OF A DENSITY

In order to estimate a continuous density f of a random variable Y,
the design proposed is to sample from a bivariate population $(X,Y) \sim \beta(x,y)$
where $X \sim \psi(x)$ is a suitably chosen concomitant variable such that
$f(y) = \int \beta(x,y)dx$ and $\psi(x) = \int \beta(x,y)dy$.

We first treat the case when ψ is a known density. It is possible to
conceive of situations where this may be the case. However, some of the
results obtained under this assumption will be used in treating the other
case when ψ is unknown.

1.1 THE CASE: ψ KNOWN

Let (X_i,Y_i), $i = 1,2,\ldots,n$ be a sample $\beta(x,y)$.
Define

$$\hat{f}_n(y) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{Y_i-y}{h}\right)$$

$$\hat{\psi}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i-x}{h}\right)$$

where $0 < h = h(n) \downarrow 0$ as $n \to \infty$, and K is a Borel-measurable bounded
function on the real line such that

$$\int K(u)du = 1, \quad \int u\, K(u)du = 0, \quad \int u^2\, K(u) < \infty,$$

$$\int |K(u)|du < \infty, \quad \text{and} \quad |uK(u)| \to 0 \quad \text{as} \quad |u| \to \infty$$

Throughout the remainder of this, we denote $\int K^2(u)du$ by $L_2(K)$ and $2^{-1} \int u^2 K(u)du$ by k_2.

We propose a ratio type estimator of $f(y)$. For $y \in S_f \equiv \{v: f(v) > 0\}$, define

$$\hat{f}_R(y) = \hat{f}_n(y)\psi(x)/\hat{\psi}_n(x)$$

where x is a suitably chosen point from S_ψ^*. The estimator $\hat{f}_R$ is well defined as it follows from Parzen (1962) that

$$P[\hat{f}_n(y)>0] \to 1, \quad \forall y \in S_f$$

and

$$P[\hat{\psi}_n(x)>0] \to 1, \quad \forall x \in S_\psi$$

as $n \to \infty$.

Let

$$\varepsilon_n = \{\hat{f}_n(y) - E\hat{f}_n(y)\}\{E\hat{f}_n(y)\}^{-1}$$

and

$$\delta_n = \{\hat{\psi}_n(x) - E\hat{\psi}_n(x)\}\{E\hat{\psi}(x)\}^{-1}$$

* The estimators $\hat{f}_n(y)$ and $\hat{\psi}_n(x)$ are standrad non-parametric estimators of the respective densities f and ψ based on a technique proposed by Rosenblatt (1956) and later extended by Parzen (1962) to the now familiar Kernel method of estimation.

Now, for the rest of this section, we will assume that

$$y \in C_f \equiv S_f \bigcap \{y: \ f''(y) \ \text{Continuous}\}$$

and

$$y \in C_\psi.$$

For simplicity, we will not write the argument of any function, i.e., we will write ϕ to denote $\phi(\cdot)$. Momentarily, we will drop the subscript n from $\hat{f}_n$, $\hat{\psi}_n$, ϵ_n and δ_n.

In terms of the r.v.'s ϵ and δ, we have

$$\hat{f}_R = \psi \cdot E\hat{f} \cdot (E\hat{\psi})^{-1}\{(1+\epsilon)(1+\delta)^{-1}\}$$

and

$$\hat{f}_R^2 = \psi^2 \cdot (E\hat{f})^2(E\hat{\psi})^{-2}\{(1+\epsilon)^2(1+\delta)^{-2}\}$$

Ignoring the terms of the order $O(nh)^{-3/2}$ and lower (See Remark 1.1 below) in the expressions for $E\hat{f}_R$ and $E\hat{f}_R^1$, we can write

$$E\hat{f}_R = \psi \cdot E\hat{f} \cdot (E\hat{\psi})^{-1}\{1-E(\epsilon\delta)+E(\delta^2)\}$$

$$(1.1)$$

$$E\hat{f}_R^2 = \psi^2(E\hat{f})^2(E\hat{\psi})^{-2}\{1-E\epsilon^2-\psi E(\epsilon\delta)+3 \ E \ \delta^2\}$$

This gives, again ignoring the terms of the order $O(nh)^{-3/2}$ and lower, an approximation for the variance

$$\sigma^2(\hat{f}_R) = \psi^2(E\hat{f})^2(E\hat{\psi})^{-2}\{E\delta^2 - 2E(\epsilon\delta) + E\epsilon^2\} \qquad (1.2)$$

$$= \psi^2(E\hat{\psi})^{-2}\sigma^2(\hat{f})$$

$$+ \psi^2(E\hat{f}/E\hat{\psi})^2\{E\delta^2 - 2 \; E(\epsilon\delta)\}$$

1.1   REMARK

In the approximation (1.1) of $E\hat{f}_R$, the error of approximation, in absolute value, is less than $E|(1+\epsilon)(1+\delta)-1(-\delta^3)|$ and

$$E|\frac{1+\epsilon}{1+\delta} \; \delta^3| \; \leq \; \{E \; \frac{1}{(1+\delta)^2}\}^{1/2}\{E \; (1+\epsilon)^2\delta^6\}^{1/2}$$

where

$$1+\delta = \hat{\psi}/E\hat{\psi}_n = \bar{w}(\text{say})$$

follows, as $n \to \infty$, a normal distribution with mean 1 and variance $\sigma^2[\bar{w}] = O(nh)^{-1}$, in fact

$$\sigma^2[\bar{w}] \sim (nh)^{-1} \; \frac{\psi(x)}{[E\psi_n]^2} \; \int K^2(u)du$$

[see Parzen (1962)].

Consequently,

$$E(\frac{1}{\bar{w}^2}) = E(\frac{1}{\bar{w}^2} I_{|\bar{w}-1| \, > \, \tau \, \log nh/\sqrt{nh}})$$

$$+ E(\frac{1}{\bar{w}^2} I_{|\bar{w}-1| \, \leq \, \tau \, \log nh/\sqrt{nh}})$$

$$\leq (1+\tau \, \log nh/\sqrt{nh})^{-2} \, P \, (\sqrt{nh} \, |\bar{w}-1| \, > \, \tau \, \log nh)$$

$$+ (1-\tau \, \log nh/\sqrt{nh})^{-2} \, P \, (\sqrt{nh} \, |\bar{w}-1| \, < \, \tau \, \log nh)$$

$$= 1 + o(1)$$

Further, since $(1+\epsilon) \stackrel{\mathcal{D}}{=} N(1 , O(nh)^{-1})$ and $\delta \stackrel{\mathcal{D}}{=} N(0 , O(Nh)^{-1})$, it follows that

$$\left(E[(1+\epsilon)^2 \, \delta^6]\right)^{1/2} = O(nh)^{-3/2}$$

Similarly, it can be shown that the error of approximation (1.1) of $E(\hat{f}_R)^2$ is $O(nh)^{-3/2}$.

It follows from Singh (1977) that

$$E\hat{f} = f + f'' k_2 h^2 + 0(h^2)$$

$$E\hat{\psi} = \psi + \psi'' k_2 h^2 + 0(h^2)$$

$$\sigma^2(\hat{f}) = (nh^2) \left[EK^2 h^{-1}(Y_1-1) - E^2 K h^{-1}(Y_1-1) \right] \qquad (1.3)$$

$$= (nh)^{-1} f L_2 (K)$$

$$+ n^{-1} \{f' \int v K^2(v) dv - f^2\} + 0(n^{-1}h)$$

$$\sigma^2(\hat{\psi}) = (nh)^{-1} \psi L_2(k)$$

$$+ n^{-1} \{\psi' \int u k^2(u) du - \psi^2\} + 0(n^{-1}h)$$

Further, since

$$\text{Cov}(\hat{f},\hat{\psi}) = (nh)^{-2} \sum_{i=1}^{n} \text{Cov}\left(K(h^{-1}(Y_i-y)), K(h^{-1}(X_i-x))\right)$$

$$= n^{-1} \left\{ \iint k(u)K(v)\beta(x+hu,y+hv) du\ dv \right.$$

$$\left. - \left(\int k(u)\psi(x+hu)\right) \left(\int K(v)f(y+hv)\right) \right\} \qquad (1.4)$$

$$= n^{-1}\{\beta-\psi f\} + 0(n^{-1}h)$$

provided first order partial derivatives of β are continuous at (x,y).

Now we will prove our main theorem:

1.1 THEOREM

For $\forall\, y \in C_f$ and $\forall\, x \in C_\psi$ such that the first order partial derivatives of β are continuous at (x,y),

$$E\hat{f}_R = f + h^2 k_2(f' - \psi''/\psi) + O(h^2)$$

$$\sigma^2(\hat{f}_R) = \sigma^2(\hat{f}) + A + O(n^{-1}h) \qquad (1.5)$$

where

$$A = n^{-1}f^2 \left\{ -2(f\psi)^{-1}(\beta - f\psi) \right.$$
$$\left. + \psi^{-2}\left(\psi'\int uK^2(u)du - \psi^2\right) + (h\psi)^{-1}\int K^2(u)du \right\}$$

and

$\sigma^2(\hat{f})$ as given in (1.3).

PROOF. From (1.3) and (1.4)

$$E\varepsilon^2 = \sigma^2(\hat{f})(E\hat{f})^{-2}$$

$$= f^{-2}\left\{(nh)^{-1}f\int K^2(u)du \right.$$

$$\left. + n^{-1}(f'\int uK^2(u)du - f^2) + O(n^{-1}h)\right\}$$

$E(\delta^2)$ has a similar expression with f being replaced by ψ.

and

$$E(\epsilon\delta) = (f\psi)^{-1} n^{-1}\{\beta-f\psi\} + O(n^{-1}h).$$

Now (1.2) and (1.3) followed by the expressions given for $E\epsilon^2$, $E\delta^2$ and $E(\epsilon\delta)$ complete the proof of the theorem.

The following corollary is an immediate consequence of Theorem 1.1

1.1 COROLLARY

If $\int uK^2(u) = 0$ (e.g., uniform or standard normal kernel) then

$$\sigma^2(\hat{f}) = (nh)^{-1}f \int K^2 - n^{-1} f^2 + O(n^{-1}),$$

and

$$\sigma^2(\hat{f}_R) = (nh)^{-1} \quad f+(f^2/\psi) \int K^2 - 2n^{-1}(1+ \frac{\beta-f\psi}{f\psi}) + o(n^{-1})$$

1.2 REMARK (COMPARISON OF $\hat{f}_R$ WITH THE USUAL ESTIMATOR $\hat{f}$)

Under the similar conditions, $E\hat{f} = f + h^2 k_2 f'' + o(h^2)$. Comparing this with the $E\hat{f}_R$ in (1.5) we see that $|Bias(\hat{f}_R)| < |Bias(\hat{f})|$ if and only if $0 < (\psi''/f'') < 2\psi$. For example, with $f(t) = \psi(t) = (2\pi)^{-1/2}exp(-t^2/2)$, this condition is satisfied if
$x^2 < 1 + 2(y^2-1)f(y)$ for $|y| \geq 1$, and if $x^2 \geq 2(1-y^2)f(y)$ for $|y| \leq 1$.

Comparing the variances of $\hat{f}_R$ and $\hat{f}$, we see that $\sigma^2(\hat{f}_R) < \sigma^2(\hat{f})$ if and only if

$$h^{-1} \int K^2 < 2 \frac{\beta - \psi f}{f} + \psi \qquad (1.7)$$

Thus, if concomitant variable is chosen in such a way that X and Y have positive dependence (i.e., $P[X \leq x, Y \leq y] \geq P[X \leq x] \, P[Y \leq y]$), all we need is to choose x and K such that $h^{-1} \int K^2 < \psi(x)$.

If $\left(2\psi(x|y) - \psi(x) \right) \geq C_0(x,y) > 0$, where $\psi(x|y)$ is the conditional density of X at $X = x$ given $Y = y$, then we can always satisfy (1.7) by choosing x and K to make $h^{-1} \int K^2 \leq C_0(x,y)$. Since the choice of X is at our will, for a given y it may be possible to include a concomitant variable X in our design and to choose x such that $2\psi(x|y) > \psi$.

1.2 THEOREM ASYMPTOTIC NORMALITY

If $h^2 = o(nh)^{-1/2}$, then

$$(nh)^{1/2} (\hat{f}_R - f) \xrightarrow{D} N(0, f^2\{f^{-1} + \psi^{-1}\} \int K^2).$$

PROOF.

Since $E(\epsilon\delta) = O(n^{-1})$ and $\sigma^2(\delta) = O(nh)^{-1}$, we write

$$\hat{f}_R = (E\hat{f})(E\hat{\psi})^{-1} \, \psi \, [1 + \epsilon - \delta + O_p(nh)^{-1}].$$

Therefore,

$$(nh)^{1/2} (\hat{f}_R - f) = (nh)^{1/2} \{(E\hat{f})(E\hat{\psi})^{-1}\psi - f\}$$

$$+ (nh)^{1/2} (E\hat{f})(E\hat{\psi})^{-1} \, \psi \, (\epsilon - \delta) + O_p(nh)^{-2} \tag{1.8}$$

From (1.3), $\left(E\hat{f} \cdot (E\hat{\psi})^{-1} \cdot \psi - f\right) = O(h^2)$, the first time on the right hand side of (1.8) is $o(1)$. Further, since $(nh)^{1/2} (\hat{f} - E\hat{f}) \xrightarrow{D} N(0, f \int K^2)$ and $(nh)^{1/2}(\hat{\psi} - \psi) \xrightarrow{D} N(0, \psi \int K^2)$ by Parzen (1962), and $Cov(\epsilon, \delta) = O(n^{-1})$, we conclude that the second term of the rhs of (1.8) is asymptotically normal with mean zero and variance $(f^2)[(f)^{-1} + (\psi)^{-1}] \int K^2$.

The proof of the theorem is now complete.

1.3   REMARK

In computing the asymptotic variance of $(nh)^{1/2} \hat{f}_R$ we have ignored the terms of the order $O(h)$ and lower, and hence the asymptotic variance of $(nh)^{1/2} \hat{f}_R$ turns out to be larger than $f \int K^2$, the asymptotic variance of $(nh)^{1/2} \hat{f}$. We have, however, seen through the proof of Theorem 1.1 that if we retain the terms of order $O(h)$ in computing the variance of $(nh)^{1/2} \hat{f}_R$, then there exist situations where $\hat{f}_R$ has smaller variance and MSE compared to those of the usual estimator $\hat{f}$.

1.2   THE CASE OF UNKNOWN ψ.

Since the choice of the concomitant variate X is at our will, we choose here that concomitant variate X which is extremely cheap to measure compared to Y variate so that we can have a very large sample on X with very little extra budget. For example, if Y is some biochemical content in a plant and X is chosen as the weight of the plant, the above condition is satisfied.

Let β denote the joint p d f of (X,Y) so that $f(y) = \int \beta(x,y)dx$ and $\psi(x) = \int \beta(x,y)dy$. Let $Z_1, \ldots, Z_{n_a}$ be n_a additional i.i.d. observations on X, independent of the paired data $(X_1, Y_1), \ldots, (X_n, Y_n)$ ~ i.i.d. according to β. We take n_a large enough so that $(n_a h_a)^{-1} = o(n^{-1})$ where $h_a = h(n_a)$. Define

$$\tilde{\psi}(x) = (n_a h_a)^{-1} \sum_{j=1}^{n_a} K(\frac{Z_j - x}{h_a}).$$

Our proposed estimator of $f(y)$ is

$$\tilde{f}_R(y) = \frac{\hat{f}(y)}{\hat{\psi}(x)} \cdot \tilde{\psi}(x)$$

where $y \in S_f$ and $x \in S_\psi$. For the sake of simplicity, we will again not display the arguments in functions like $\tilde{f}_R(y)$, $\hat{f}(y)$, etc.

Since $E\tilde{\psi} = \psi + 0(h_a^2)$, it follows from subsection 1.1 that

$$E\tilde{f}_R(y) = E\left(\frac{\hat{f}(y)}{\hat{\psi}(x)}\right) \cdot E(\tilde{\psi}) = f + 0(h^2).$$

Now we examine the variance of $\tilde{f}_R$. Since for independent random variables W and V,

$$\sigma^2(WV) = EW^2 \cdot EV^2 - E^2W \cdot E^2V$$

$$= \sigma^2(W)\sigma^2(V) + E^2(W)\sigma^2(V) + E^2(V)\sigma^2(W),$$

we can write with $\hat{f}_R$ as given in subsection 1.1,

$$\sigma^2(\tilde{f}_R) = \sigma^2\left(\frac{\hat{f}}{\hat{\psi}}\right)\sigma^2(\tilde{\psi}) + E^2\left(\frac{\hat{f}}{\hat{\psi}}\right)\sigma^2(\tilde{\psi}) + E^2(\tilde{\psi})\,\sigma^2\left(\frac{\hat{f}}{\hat{\psi}}\right)$$

$$= \sigma^2(\hat{f}_R)\,\frac{\sigma^2(\tilde{\psi})}{\psi^2} + (E\hat{f}_R)^2\,\frac{\sigma^2(\tilde{\psi})}{\psi^2} + \frac{(E\tilde{\psi})^2}{\psi^2} \cdot \sigma^2(\hat{f}_R)$$

$$= 0\left((nh)^{-1}(n_a h_a)^{-1}\right) + 0(n_a h_a)^{-1} + \left(1 + 0(h_a^2)\right)\sigma^2(\hat{f}_R)$$

Since

$$\sigma^2(\hat{f}_R) = O(nh)^{-1} \; , \; \sigma^2(\tilde{\psi}) = O(n_a h_a)^{-1} \; \text{ and } \; E\hat{f}_R = f + O(h^2)$$

Thus, since $(n_a h_a)^{-1} = o(n^{-1})$, we have

$$\sigma^2(\tilde{f}_R) = \sigma^2(\hat{f}_R) + o(n^{-1}) = \sigma^2(\hat{f}) + A + o(n^{-1})$$

where A is as given in Theorem 1.1. Therefore, the conclusion of Remark 1.1 continues to hold.

With regard to the asymptotic distribution of $\tilde{f}_R$, we note that

$$\tilde{\psi} = \psi + O_p(n_a h_a)^{-1/2} = \psi + o_p(n^{-1/2}).$$

Hence

$$\tilde{f}_R = \hat{f}_R + o_p(n^{-1/2})$$

and

$$(nh)^{1/2}(\tilde{f}_R - f) \xrightarrow{D} (nh)^{1/2}(\hat{f}_R - f) \xrightarrow{D} N\left(0 \; , \; f^2\{f^{-1} + \psi^{-1}\}\int K^2\right)$$

from Theorem 1.2

1.4 REMARK REGRESSION TYPE DENSITY ESTIMATORS

We propose a linear regression type density estimator of f as

$$f_{1r}(y) = \hat{f}(y) - b\left(\hat{\psi}(x) - \psi(x)\right) .$$

Then

$$E\hat{f}_{1r} = f + h^2 h_2\left(f''(y) - b\psi''(x)\right) + o(h^2),$$

and

$$\sigma^2(\hat{f}_{1r}) = \sigma^2\left(\hat{f}(y)\right) + b^2\sigma^2\left(\hat{\psi}(x)\right) - 2b \; \text{Cov}\left(\hat{f}(y) \; , \; \hat{\psi}(x)\right)$$

where $\sigma^2(\hat{f})$, $\sigma^2(\hat{\psi})$ and $\text{Cov}(\hat{f} \; , \hat{\psi})$ are as given in (1.3) and (1.4). Thus

$$\sigma^2(\hat{f}_{1r}) < \sigma^2\left(\hat{f}(y)\right)$$

if and only if

$$b^2 \leq 2b \; \frac{\text{Cov}\left(\hat{f}(y) \; , \; \hat{\psi}(x)\right)}{\sigma^2\left(\hat{\psi}(x)\right)}$$

$$= 2b \; \frac{(\beta - \psi f)}{h^{-1}\psi \int K^2 - \psi^2}$$

2. ESTIMATION OF A CONDITIONAL DENSITY

Let $g(y|x) = \beta(x,y)/f(x)$ be the conditional density of $Y|X = x$, where the couple $(X,Y) \sim \beta(x,y)$, $X \sim f(x) = \int \beta(x,y)dy$ and $f(x) > o$.

Rosenblatt (1969) treats the problem of estimating g on the basis of a random sample $(X_1,Y_1),\dots,(X_{n_c},Y_{n_c})$ from the joint distribution of (X,Y). We are also going to estimate g but under a data set up which is slightly more general. In addition to n_c paired observations (X_i,Y_i)'s we also have additional data on X, i.e., a sample from the univariate distribution of X,

$$U_1, U_2,\dots, U_{n_a}.$$

Set

$$N = n_c + n_a$$

and

$$Z_j \begin{cases} = X_j & \text{for } j = 1,2,\ldots,nc \\[2em] = U_{j-n_c} & \text{for } j = n_c+1,\ldots,N \end{cases}$$

Let $h(t)$ be a positive function such that

$$h(t) \downarrow 0 \quad \text{and} \quad th^2(t) \uparrow \infty$$

as $t \to \infty$. Set

$$h_c = h(n_c),$$

$$h = h(n_c+n_a)$$

and note that, as $n_c \uparrow \infty$,

$$h_c \downarrow 0 \, , \quad h \downarrow 0 \, , \quad n_c h_c^2 \uparrow \infty \quad \text{and} \quad Nh \uparrow \infty.$$

Further, let $B(u,v)$ be a Borel measurable bounded function defined on R^2 such that as

$$||(u,v)|| \to \infty \, , \quad ||(u,v)|| \; |B(u,v)| \to 0.$$

We also assume that

$$\iint |B(u,v)|du \, dv < \infty,$$

$$\iint B(u,v)du \, dv = 1,$$

and

$$\iint uB(u,v)du\ dv = 0 = \iint vB(u,v)du\ dv$$

Also, let K be Borel-measurable bounded function defined on the real line such that

$$\frac{\lim}{u \to \infty}\ |uK(u)| = 0,$$

$$\int |K(u)|du < \infty\ ,\quad \int K(u)du = 1\ ,\quad \int uK(u) = 0$$

and

$$\int u^2 K(u)du < \infty.$$

Having chosen the weight functions B and K and the sequence of bandwidths {h(n)}, we propose the following estimator for $g(y|x)$ at a point of continuity (x,y) of $\beta(x,y)$ such that $f(x) > 0$.

Define

$$\hat{g}_{AS}(y|x) = \hat{\beta}_{n_c}(x,y)\ /\ \hat{f}_N(x)$$

with

$$\hat{f}_N(x) = (Nh)^{-1} \sum_{j=1}^{N} K\left(h^{-1}(Z_j-x)\right)$$

and

$$\hat{\beta}_{n_c}(x,y) = (n_c h_c^2)^{-1} \sum_{j=1}^{n_c} B\left(h_c^{-1}(X_j-x),\ h_c^{-1}(Y_j-y)\right)$$

2.1 REMARK

It follows from Parzen (1962), that if $f(x) > o$, then
$P[\hat{f}_N(x) > 0] \to 1$ and $P[\hat{f}_{n_c}(x) > o] \to 1$. Therefore, $\hat{g}_{AS}$ is well-defined
in probability.

2.2 REMARK

When there is no additional data, i.e., the case when $\dot{n}_a = o$, $\hat{g}_{AS}$
reduces to the estimator studied by Rosenblatt (1969);

$$\hat{g}(y|x) = \hat{\beta}_{n_c}(x,y) / \hat{f}_{n_c}(x)$$

where

$$\hat{f}_{n_c}(x) = (\dot{n}_c h_c)^{-1} \sum_{j=1}^{n_c} K\left(h_c^{-1}(X_j - x)\right)$$

It is well known, (e.g., Rosenblatt (1956) and Cacoullos (1966)), that
$\hat{f}_N(x)$ as an estimator of $f(x)$ and $\hat{\beta}_{n_c}(x,y)$ as that of $\beta(x,y)$ are
consistent in quadratic mean. Intuitively, we expect $\hat{g}_{AS}$ to estimate g
consistently. We prove this and other results in the remainder of this
section.

As before, for the remainder of Section 2, we will not display the
arguments in the functions defined above.

2.1 ASYMPTOTIC APPROXIMATIONS FOR BIAS, VARIANCE, AND THE DISTRIBUTION OF $\hat{g}_{AS}(y|x)$.

In this section we show that $\hat{g}_{AS}$ as an estimator of g is asymptotically unbiased, consistent in quadratic mean and asymptotically normal just like the usual estimators of $g(y|x)$ proposed by Rosenblatt (1969), which are based on only paired observations. Approximation for the bias and variance obtained here for $\hat{g}_{AS}$, specialized to the Rosenblatt's case (i.e., when $n_a = 0$), are better than those noted in Rosenblatt (1969). We further give sufficient conditions on β and f under which the absolute bias and variance of $\hat{g}_{AS}$ are smaller than those for $\hat{g}$ obtained by Rosenblatt.

Although we have investigated the asymptotic properties with $n_a \to \infty$, we have observed (though not reported here), through Monte-Carlo simulation that for n_c fixed the estimators $\hat{g}_{AS}(y|x)$ proposed here have in some cases smaller mean squared error than the usual estimators.

It is well known (e.g., Singh (1977)), that if f'', the second derivative of f, is continuous at x, then with $1_1(x) = f''(x)\int u^2 K(u)du / 2$ and $L_2(K) = \int K^2(u)du$, we have

$$E\hat{f}_N = f + h^2 1_1 + o(h^2)$$

$$E \hat{f}_{n_c} = f + h_c^2 1_1 + o(h_c^2)$$

$$\sigma^2(\hat{f}_N) = (Nh)^{-1} f L_2(K) + o(Nh)^{-1}$$

$$\sigma^2(\hat{f}_{n_c}) = (n_c h_c)^{-1} f L_2(K) + o(n_c h_c)^{-1},$$

(2.0)

and with

$$l_2(x,y) = \frac{\partial^2 \beta(x,y)}{2\partial x \partial y} \iint u^2 B(u,v) + \frac{\partial^2 \beta(x,y)}{\partial y^2} \iint v^2 B(u,v) du\ dv$$

and

$$L_2(B) = \iint B^2(u,v) du\ dv,$$

Choosing B in such a way that $\iint u B(u,v) du\ dv = 0$

$= \iint v B(u,v) du\ dv$, we obtain from Rosenblatt (1969) and the techniques used in Theorem 1 of Parzen (1962) that

$$E\hat{\beta}_{n_c} = \beta + h_c^2 l_2 + o(h_c^2),$$

$$\sigma^2(\hat{\beta}_{n_c}) = (n_c h_c^2)^{-1} \beta L_2(B) + (n_c h_c)^{-1} \frac{\partial \beta}{\partial x} \cdot \iint u B^2(u,v) du\ dv$$

$$+ \frac{\partial \beta}{\partial y} \cdot \iint v B^2(u,v) du\ dv + o(n_c h_c)^{-1}$$

$$= (n_c h_c^2)^{-1} \beta L_2(B) + o(n_c h_c)^{-1}$$

For the rest of this section, put $\gamma_1(x) = l_1(x) / f(x)$ and $\gamma_2(x,y) = \{l_2(x,y) / \beta(x,y) - l_1(x) / f(x)\}$. As with others, the functions γ_1 and γ_2 will be displayed without their arguments.

Let

$$\epsilon = (\hat{\beta}_{n_c} - E\hat{\beta}_{n_c})(E\hat{\beta}_{n_c})^{-1}$$

and

$$\delta = (\hat{f}_N - E\hat{f}_N)(E\hat{f}_N)^{-1}$$

Then, in terms of ϵ and δ, we have

$$\hat{g}_{AS} = (E\hat{\beta}_{n_c})(E\hat{f}_N)^{-1} \{(1 + \epsilon)(1 + \delta)^{-1}\}. \tag{2.2}$$

It is well known that ϵ and δ are $0_p(n_c h_c)^{-1/2}$ and $0_p(Nh)^{-1/2}$ respectively. Further, these are asymptotically normal random variables with mean zero and with their variances tending towards zero as $n_c h_c^2 \to \infty$ in case of ϵ and as $Nh \to \infty$ in case of δ. Therefore, it follows that

$$\hat{g}_{AS} = (E\hat{\beta}_{n_c})(E\hat{f}_N)^{-1} \{1 + \epsilon - \delta - \epsilon\delta + \delta^2\} + 0_p(n_c h)^{-3/2}. \tag{2.3}$$

Further in view of the comments made in Remark 1.1 of Section 1, it follows that

$$E\hat{g}_{AS} = (E\hat{\beta}_{n_c})(E\hat{f}_N)^{-1} \{1 - E(\epsilon\delta) + E(\delta^2)\} + 0(n_c h)^{-3/2}$$

and

$$\tag{2.4}$$

$$E(\hat{g}_{AS}^2) = (E\hat{\beta}_{n_c})^2(E\hat{f}_N)^{-2} \{1 + E\delta^2 - 4E(\epsilon\delta) + 3E\delta^2\} + 0(n_c h)^{-3/2}$$

With the above observations, we are now able to prove asymptotic unbiasedness, quadratic mean consistency, and the asymptotic normality of $\hat{g}_{AS}$. Throughout the remainder of this section, we assume that $h_c = h(n_c)$ is such that $\lambda_{n_c} = h_c/h \to \lambda < \infty$ and K is such that

$K(\lambda_c u) \to K(\lambda u)$ a.e., in u as $n \to \infty$ (this is assured if K is continuous a.e.). To prove our main results, we make use of the following lemma.

2.1 LEMMA

If β is continuous at (x,y), then

$$\text{Cov}(\hat{\beta}_{n_c}, \hat{f}_N) = (Nh)^{-1}\beta L_\lambda(KB) + o(Nh)^{-1}$$

where

$$L_\lambda(KB) = \iint B(u,v) K(\lambda u) du\, dv.$$

PROOF.

Since (X_j, Y_j), $j = 1,\ldots,n_c$ are i.i.d. and are independent of $\{U_j, j = 1,\ldots n_a\}$,

$$\text{Cov}(\hat{\beta}_{n_c}, \hat{f}_N) = (n_c h_c^2 Nh)^{-1} \sum_{j=1}^{n_c} \text{Cov}\left(B\left(\frac{x-X_j}{h_c}, \frac{y-Y_j}{h_c}\right), K\left(\frac{x-X_j}{h}\right)\right)$$

$$= (Nh)^{-1} [A_{n_c} - A'_{n_c}]$$

where

$$A'_{n_c} = h_c^{-2} \left[\dot{B}\left(\frac{x-X_1}{h_c}, \frac{y-Y_1}{h_c}\right) K\left(\frac{x-X_1}{h}\right)\right]$$

and

$$A_{n_c} = h_c^{-2} E B\left(\frac{x-X_1}{h_c}, \frac{y-Y_1}{h_c}\right) \cdot EK\left(\frac{x-X_1}{h}\right).$$

Now consider first A_{n_c}. We can write

$$|A_{n_c} - \beta L_\lambda(KB)| \le \gamma_1(x,y,n_c) + \gamma_2(x,y,n_c)$$

where

$$\gamma_1(x,y,n_c) = \left|\iint \left(\beta(x - h_c u, y - h_c v) - \beta(x,y)\right) K(\lambda_c u)B(u,v) du\, dv\right|$$

and

$$\gamma_2(x,y,n_c) = \beta(x,y) \left| \iint \{K(\lambda_{n_c} u) - K(u)\} B(u,v) du\, dv \right|.$$

Since K is bounded, it follows from Cacoullos (1966) that $\gamma_1 = o(1)$, and since $K(\lambda_{n_c} u) \to K(\lambda u)$, by dominated convergence theorem, γ_2 is also $o(1)$. Hence, $A_{n_c} = \beta L_\lambda(KB) + o(1)$.

Further, from Cacoullos (1966),

$$EB\left(\frac{x-X_1}{h_c}, \frac{y-Y_1}{h_c}\right) = h_c^2 \beta(x,y) + o(h_c^2)$$

and

$$EK\left(\frac{x-X_1}{h}\right) = hf(x) + o(h).$$

Therefore, $A'_{n_c} = h\beta f + o(h)$.

The proof of the lemma is now complete.

2.1 THEOREM ASYMPTOTIC UNBIASEDNESS

If the second order partial derivatives of β are continuous at (x,y), then

$$E\left(\hat{g}_{AS}(y|x) - g(y|x)\right) = g(y|x)[h_c^2 \gamma_2(x,y) + (h_c^2 - h^2)\, \gamma_1(x)$$

$$+ (Nh)^{-1} \frac{g(y|x)}{f(x)} \{L_2(K) - L_\lambda(BK)\} \qquad (2.6)$$

$$+ o(\max \{h_c^2, (Nh)^{-1}\})].$$

PROOF.

It follows from (2.0) and Lemma 2.1 that

$$E(\epsilon \cdot \delta) = (E\hat{\beta}_{n_c})^{-1}(E\hat{f}_N)^{-1} \text{Cov}(\hat{\beta}_{n_c}, \hat{f}_N)$$

$$= (Nh)^{-1} f(x)^{-1} L_\lambda(KB) + o(Nh)^{-1} \qquad (2.7)$$

This result accompanied by (2.4) and (2.0) gives

$$E\hat{g}_{AS} = g\left[\left(1 + \frac{l_2 h_c^2}{\beta} + o(h_c^2)\right) \left(1 + \frac{l_1 h^2}{f} + o(h^2)\right)^{-1} \right]$$

which finally gives (2.6).

2.2 REMARK

Notice that if $(Nh)^{-1} = o(h_c^2)$, then the bias in Theorem 2.1 is given by

$$E\left(\hat{g}_{AS}(y|x) - g(y|x)\right) = g(y|x) \left\{ \frac{l_2(x,y)h_c^2}{\beta(x,y)} - \frac{l_1(x)h^2}{f(x,y)} \right\} + o(h_c^2). \qquad (2.6)'$$

The right hand side of this equation with $n_a = 0$ reduces to what Rosenblatt (1969) has noted for the bias of the estimator $\hat{g}$. Writing (2.6)' as

$$E\left(\hat{g}_{AS}(y|x) - g(y|x)\right) = g(y|x) h_c^2 \gamma_2(x,y)$$
$$+ g(y|x)(h_c^2 - h^2) \gamma_1(x) + o(h_c^2) \qquad (2.6)''$$

we see that the first term on the right hand side of (2.6)" is the bias of $\hat{g}$ with no additional data.

Thus, we conclude the following corollary:

2.1 COROLLARY

Let $(Nh)^{-1} = o(h_c^2)$. Under the hypothesis of Theorem 2.1,

$$|\text{bias of } \hat{g}_{AS}(y|x)| < |\text{bias of } \hat{g}(y|x)|$$

if and only if $\gamma_2(x,y)$ and $\gamma_1(x)$ are of opposite signs and

$$\left(1 - \frac{h^2}{h_c^2}\right) |\gamma_1(x)| < 2|\gamma_2(x,y)|.$$

2.2 THEOREM   VARIANCE OF $\hat{g}_{AS}$

If β is continuous at (x,y), then

$$\sigma^2\left(\hat{g}_{AS}(y|x)\right) = g(y|x)\left(f(x)\right)^{-1}[(Nh)^{-1}g(y|x)\{L_2(K)-2L_\lambda(KB)\}$$
$$+ (n_c h_c^2)^{-1} L_2(B)] + o(\max\{(Nh)^{-1}, (n_c h_c)^{-1}\}). \tag{2.8}$$

PROOF.

It follows from (2.4) that

$$\sigma^2\left(\hat{g}_{AS}(y|x)\right) = (E\hat{\beta}_{n_c})^2(E\hat{f}_N)^{-2}\{E\delta^2 - 2E\epsilon\delta + E\epsilon^2\} + O(n_c h)^{-3/2}$$

In view of (2.0), (2.1), and (2.7), the right hand side is

$$g^2(y|x)\left[(Nh)^{-1}\left(f(x)\right)^{-1}\{L_2(K) - 2L_\lambda(KB)\} + o(Nh)^{-1}\right.$$
$$\left. + (n_c h_c^2)^{-1}\left(\beta(x,y)\right)^{-1} L_2(B) + o(n_c h_c)^{-1}\right]$$

which is the right hand side of (2.8).

This completes the proof of the theorem.

2.3 REMARK

The estimator proposed by Rosenblatt (1969), which is only based on a set of paired observation, coincides with our estimator in the case $n_a = 0$. However, his approximation to the variance of $\hat{g}(y|x)$ is

$$(n_c h_c^2)^{-1} \, g(y|x) \, L_2(B) \, / \, f(x) + o(n_c h)^{-1}$$

which is strictly larger than the approximation obtained by us. For the case $n_a = 0$, our approximation for the variance of $\hat{g}(y|x)$ is

$$\left\{ (n_c h_c^2)^{-1} \, g(y|x) \, L_2(B) - (n_c h)^{-1} \, g^2(y|x) \, L_2(K) \right\} \, / \, f(x) + o(n_c h_c)^{-1}$$

The following corollary is a consequence of Theorem 2.1 and 2.2.

2.1 COROLLARY QUATRATIC MEAN CONSISTENCY OF $\hat{g}$

Under the conditions of Theorem 2,

$$
\begin{aligned}
\text{MSE} &\left[\hat{g}_{AS} \right] \\
&= g^2 \left[\{(\beta)^{-1} l_2 \, h_c^2 - (f)^{-1} \, l_1 \, h^2 \} \right] \\
&\quad + (Nh)^{-1}(f)^{-1} \{ L_2(K) - 2L_\lambda(KB) \} \\
&\quad + (n_c h_c^2)^{-1}(\beta)^{-1} \, L_2(B) \\
&\quad + o(\max \{ (Nh)^{-1} , (n_c h_c)^{-1} \})
\end{aligned}
\tag{2.9}
$$

2.4 REMARK

If $o\left(\max\{(Nh)^{-1}, (n_c h_c)^{-1}\}\right)$ is ignored,

then

$$MSE\left[\hat{g}_{AS}\right] = W_1 + W_2$$

where

$$W_1 = g^2\left[h_c^4 \gamma_2^2 + (n_c h_c^2)^{-1}(\beta)^{-1} L_2(B)\right]$$

and

$$W_2 = (h_c^2 - h^2)\left[(h_c^2 - h^2)\gamma_1^2 + 2h_c^2 \gamma_1 \cdot \gamma_2\right]$$

When $\pi_a' = o$, the case of no additional data, $W_2 = 0$ and W_1 is the MSE of $\hat{g}(y|x)$, as is also noted by Rosenblatt (1969). Thus, we have the following corollary.

2.2 COROLLARY

If $o\left(\max\{(Nh)^{-1}, (n_c h_c)^{-1}\}\right)$ is ignored, then under the hypothesis of Theorem 2.1

$$MSE\left[\hat{g}_{AS}(y|x)\right] < \left[MSE\,\hat{g}(y|x)\right]$$

if γ_1 and γ_2 are of opposite signs, and

$$(1 - h^2/h_c^2)\,|\gamma_1(x)| < 2|\gamma_2(x,y)|. \tag{2.10}$$

The conditions stated in the corollary 2.2, under which one would recommend the use of additional data, are not of practical utility. They need to be examined more critically. Our conjecture is that $\hat{g}_{AS}$ will not perform better than $\hat{g}$ in the case of strongly dependent variables X and Y.

2.3 THEOREM ASYMPTOTIC NORMALITY OF $\hat{g}_{AS}$

If $(n'_c h_c^2)^{1/2} = o(h_c^2)$ and $(n_c N^{-1} h_c^2 h^{-1})^{1/2} = o(1)$,

then

$$(n_c h_c^2)^{-1} \left(\hat{g}_{AS}(y|x) - g(y|x) \right) \xrightarrow{\ \mathcal{D}\ } N \left(0 \ , \ g(y|x) \ \left(f(x) \right)^{-1} \Big|_B^2 \right) \qquad (2.11)$$

PROOF.

From our foregoing analysis, it follows that

$$\delta = 0_p(Nh)^{-1/2} \ , \ \varepsilon\delta = 0_p(Nh)^{-1} \ \ \text{and} \ \ \delta^2 = 0_p(Nh)^{-1}.$$

Therefore, from (2.3), we can write

$$(n_c h_c^2)^{1/2} (\hat{g}_{AS} - g) = (n_c h_c^2)^{1/2} \{ E\ \hat{\beta}_{n_c} \cdot (E\hat{f}_N)^{-1} - g \}$$

$$+ (n_c h_c^2)^{1/2} \cdot E\hat{\beta}_{n_c} \cdot (E\hat{f}_N)^{-1} \cdot \varepsilon + o_p(1) \qquad (2.12)$$

In view of (2.0) and (2.1), the first term of the right hand side of (2.12) is $o(1)$. Further, since from Cacoullos (1966),

$$(n_c h_c^2)^{1/2} \left(\hat{\beta}_{n_c} - E(\hat{\beta}_{n_c}) \xrightarrow{\ \mathcal{D}\ } N \left(0 \ , \ \beta L_2(\beta) \right) \right),$$

the second term on the right hand side of (2.12) is asymptotically normal with mean zero and variance $g \cdot (f)^{-1} L_2(K)$. The proof the theorem is now complete.

ACKNOWLEDGEMENT

REFERENCES

[1] CACOULLOS, T. (1966). Estimating of a Multivariate Density. Annals Inst. Statist. Math. 18, 179-189.

[2] LIANG, W.Q. and KRISHNAIAH, P.R. (1985). Non-parametric Iterative Estimation of Multivariate Binary Density. Jour.Multivariate Anal. 16, 162-172.

[3] MURTHY, V.K. (1965). Estimation of Jumps, Reliability and Hazard Rate. Ann. Math. Statist. 36, 1032-1040.

[4] ROSENBLATT, MURRAY (1956). Remarks on some Non-parametric Estimates of a Density Function. Ann. Math. Statist. 27, 832-837.

[5] ROSENBLATT, MURRAY (1969). Conditional Probability Density and Regression Estimators. Multivariate Analysis, Vol.2, 25-31. Ed. P.R. Krishnaiah

[6] PARZEN, EMANUEL (1962). On Estimation of a Probability Density Function and Mode. Ann. Math. Statist. 33, 1065-1076.

[7] SINGH, RADHEY S. (1977a). Improvement on some known Non-parametric Uniformly Consistent Estimators of Derivates of a Density. Ann. Statist. 5, 394-399.

[8] SINGH, RADHEY S. (1977b). Applications of Estimators of a Density and its Derivatives to Certain Statistical Problems. J. Roy. Statist. Soc. Ser.B, 39, 357-363.

[9] WATSON, G.S. (1969). Density Estimation by Orthogonal Series. Ann. Math. Statist. 40, 1496-1503.

[10] WEGMAN, E.J. (1972a). Non-parametric Probability Density Estimation .I. A Summary of Available Methods. Technometrics, 14, 533-546.

[11] WEGMAN, E.J. (1972b). Non-parametric Probability Density Estimation .II. A Comparison of Density Estimation Methods. J. Statist. Comp. and Simulation 1, 225-245.

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER AFOSR-TR- 85 - 0 7 8	2. GOVT ACCESSION NO. AD-A160287	3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle) Role of Auxiliary Variate and Additional Data in Density Estimation	5. TYPE OF REPORT & PERIOD COVERED Technical - May, 1985
	6. PERFORMING ORG. REPORT NUMBER 85-21

7. AUTHOR(s) Manzoor Ahmad and Radhey S. Singh	8. CONTRACT OR GRANT NUMBER(s) F-49620-85-C-0008

9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Multivariate Analysis 515 Thackeray Hall University of Pittsburgh, Pittsburgh, PA 15260	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5

11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Department of the Air Force Bolling Air Force DC 20332	12. REPORT DATE May, 1985
	13. NUMBER OF PAGES 33

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Kernel Method; Unbiasedness; Mean Square Consistencies; Rate of Convergence.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Some new estimators of a univariate p.d.f. $f(y)$ of a random variable Y, based on a set of observations taken from a bivariate joint density $B(x,y)$ of Y and a suitably chosen concomitant variable X, have been investigated. Asymptotic unbiasedness, mean square consistency, asymptotic normality and rates of convergence have been established. A related problem of estimation of a conditional density has also been studied.

DD FORM 1473
1 JAN 73

Unclassified

END

FILMED

11-85

DTIC